

# Web-based Supplementary Materials for “Modeling and Testing for Joint Association Using a Genetic Random Field Model” by

Zihuai He<sup>1,\*</sup>, Min Zhang<sup>1,\*\*</sup>, Xiaowei Zhan<sup>1</sup> and Qing Lu<sup>2</sup>

<sup>1</sup>Department of Biostatistics, University of Michigan, Ann Arbor, U.S.A.

<sup>2</sup>Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, U.S.A.

\**email*: zihuai@umich.edu

\*\**email*: mzhangst@umich.edu

## 1. Robustness to other distributions

We evaluated the robustness of the GenRF test to distributions other than normal. The GenRF test for traits with distributions other than normal is described in Section 2.3 of the main manuscript. The simulation setup is otherwise similar to the first set of simulations, described in Section 3 of the main manuscript, with only one region,  $p = 10$ ,  $\rho = 0.4$  and  $n = 100$ . Responses  $Y_i$  were generated according to generalized linear models using the canonical link function, i.e.,

$$g(\mu_i) = aG_{i,5},$$

where  $a$  was set to be 1.1 and 2.5 respectively for exponential and binary distributions. For Mixture Normal, we generated two normal distributions with mean difference 10, equal mixture proportions, and  $a = 2.7$ . We set  $a$  to be 0 in evaluating the type-I error rate. The results are shown in Supplementary Table 1.

## 2. Robustness to heteroscedastic variances of binary traits

We evaluated the robustness of the GenRF test to heteroscedastic variances of binary traits. Since the variance of a binary outcome is a function of its mean, the variance is known to be heteroscedastic when the mean of outcome depends on covariates. The modification of the GenRF test for binary traits is described in Section 2.3 of the main manuscript. The simulation setup is otherwise similar to the first set of simulations, described in Section 3 of

Supplementary Table 1: Simulation results under different distributions of the response variable (1000 replicates). \* indicates results are unavailable due to “sample size is small, need small sample adjustment” and SKAT has no small sample adjustment for IBS kernel.

Method		Distribution		
		Exponential	Mixture Normal	Binary
GenRF	Power	0.636	0.582	0.646
	Type I	0.052	0.056	0.046
SKAT	Power	0.655	0.582	*
	Type I	0.046	0.046	*
F-test	Power	0.572	0.568	0.559
	Type I	0.056	0.054	0.050

the main manuscript, with  $p = 20$ ,  $\rho = 0$ , minor allele frequency 0.2, and  $n = 100$ . Responses  $Y_i$  were generated according to logistic models, i.e.,

$$\text{logit}(p_i) = aG_{i,5} + bX_i,$$

where  $a$  was set to be 3 in evaluating power and 0 in evaluating type I error rate;  $X_i$  was a covariate generated from  $N(0, 1)$ ; and  $b$  was varying from 0 to 10 to generate different levels of heteroscedastic variance. A larger coefficient  $b$  results in a wider range of the predicted mean and thus more heteroscedastic variance. When  $b = 5$  or 10, which represents unusually strong effect of  $X$  (probably unlikely in practice), some predicted means fall outside of  $[0, 1]$  and truncation at 0 or 1 was used. The results are shown in Supplementary Table 2. We note that the power decreases as the coefficient  $b$  increases because the noise becomes larger. The type I error is well controlled even if some predicted means reached 0 or 1, indicating that the GenRF test with the minor modification is robust to heteroscedastic variances of binary traits.

Supplementary Table 2: Simulation results under different levels of heteroscedastic variances (500 replicates). Coefficient: the coefficient of the covariate. \* indicates that some predicted means reached 0 or 1.

Method		Coefficient				
		0	1	3	5*	10*
GenRF	Power	0.538	0.594	0.408	0.290	0.110
	Type I	0.052	0.046	0.040	0.058	0.060

### 3. Application to Dallas Heart Study

We analyzed data from the Dallas Heart Study (Browning et al., 2004.), a population-based, multi-ethnic study on 3551 subjects whose lipids and glucose metabolism are measured. In this study, sequence variations in the coding regions of the four genes, ANGPTL3, ANGPTL4, ANGPTL5 and ANGPTL6 are discovered. Supplementary Table 3 lists the number of non-synonymous variants in each gene and their MAFs.

Supplementary Table 3: Dallas Heart Study sequencing data information: number of non-synonymous variants in each gene. MAF: minor allele frequency.

	Number of Variants			
	ANGPTL3	ANGPTL4	ANGPTL5	ANGPTL6
All	21	25	18	25
MAF < 5%	21	24	18	25
MAF < 1%	20	23	17	24